International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 7 Issue 2, March-April 2023 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

# Real Time Object Detection with Audio Feedback using Yolo_v3

**Dr. K. Nagi Reddy, K. Sreeja, M. Sreenivasulu Reddy, K. Sireesha, M. Triveni**

Department of ECE, N.B.K.R. Institute of Science and Technology, Tirupati District, Andhra Pradesh, India

## ABSTRACT

In this paper, we propose a system that combines real-time object detection using the YOLOv3 algorithm with audio feedback to assist visually impaired individuals in locating and identifying objects in their surroundings. The YOLOv3 algorithm is a state-of-the-art object detection algorithm that has been used in numerous studies for various applications. Audio feedback has also been studied in previous research as a useful tool for assisting visually impaired individuals. Our proposed system builds on the effectiveness of both these technologies to provide a valuable tool for improving the independence and quality of life of visually impaired individuals. We present the architecture of our proposed system, which includes a YOLOv3 model for object detection and a text-to-speech engine for providing audio feedback. We also present the results of our experiments, which demonstrate the effectiveness of our system in detecting and identifying objects in real-time. Our proposed system can be used in various settings, such as indoor and outdoor environments, and can assist visually impaired individuals in various activities such as the navigation and object identification.

***KEYWORDS:*** *Object detection, YOLO RCNN*

***How to cite this paper:*** Dr. K. Nagi Reddy | K. Sreeja | M. Sreenivasulu Reddy | K. Sireesha | M. Triveni "Real Time Object Detection with Audio Feedback using Yolo_v3" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-7 | Issue-2, April 2023, pp.857-860, URL: www.ijtsrd.com/papers/ijtsrd55158.pdf

IJTSRD55158





## I. INTRODUCTION

One of the difficult applications of computer vision is object recognition, which has been widely used in various fields, such as autonomous vehicles, robotics, security tracking, and guiding visually impaired people. Many algorithms were increasing the connection between video analysis and picture understanding as deep learning advanced quickly. Using varied network architectures, each of these techniques accomplishes the same task of multiple object detection in complicated images. The freedom of movement in an unknown environment is restricted by the absence of vision impairment, thus it is crucial to use modern technologies and teach them to assist blind people whenever necessary.

Python module used to translate statements into audio speech in order to obtain the audio Feedback gTTS (Google Text to Speech). The Python module is used to play the audio in the project. Both algorithms are examined using webcams in various scenarios to assess algorithm accuracy in every scenario.

## II. LITERATURE SURVEY:

1. C. Senthil Singh and Sherin Cherian. Implementation of object tracking in real time using a camera. Real-time object tracking and detection are crucial functions in many computer vision systems. Variations in object shape, partial and total occlusion, and scene illumination pose serious challenges for reliable object tracking. We provide a method for object detection and tracking that combines kalman filtering and Prewitt edge detection. The two key components of object tracking that can be accomplished by applying these methods are the representation of the target item and the location prediction. Real-time object tracking is created here using a webcam. Tests demonstrate that our tracking system can efficiently track moving objects even when they are deformed or obscured, as well as track several objects.
2. Shou-tao Xu, Zhong-Qiu Zhao, Peng Zheng, and Xindong Wu. Deep Learning for Object Recognition: A Review. The foundation of conventional object detection techniques is shallow trainable structures and handmade features. Building intricate ensembles that incorporate several low-level picture features with high-level context from object detectors and scene classifiers can readily stabilize their performance.

In order to solve the issues with traditional architectures, more potent tools that can learn semantic, high-level, deeper features are being offered as a result of deep learning's quick development. In terms of network architecture, training methodology, optimization function, etc., these models behave differently. In this paper, we explore object detection frameworks based on deep learning. A brief history of deep learning and its illustrative tool, the Convolutional Neural Network, is given before our review (CNN). Then, we concentrate on common generic object detection architectures with a few changes and helpful tips to further enhance detection performance. We also provide a brief overview of a number of specific tasks, such as salient item detection, face detection, and pedestrian detection, as different specific detection tasks exhibit different characteristics. Moreover, experimental studies are offered to contrast different approaches and reach some insightful results. In order to provide direction for future work in both object identification and pertinent neural network based learning systems, a number of promising directions and tasks are provided.

3. B. Triggs and N. Dalal. Oriented gradient histograms for human detection. We investigate the issue of feature sets for reliable visual object recognition using a test case of linear SVM-based human detection. In this experimental demonstration, we demonstrate that grids of histograms of oriented gradient (HOG) descriptors greatly outperform existing feature sets for human detection after examining existing edge and gradient based descriptors. We examine the impact of each computation stage on performance and come to the conclusion that fine-scale gradients, fine orientation binning, somewhat coarse spatial binning, and excellent local contrast normalization in overlapping descriptor blocks are all crucial for successful outcomes. As a result of the new method's nearly flawless separation on the original MIT pedestrian collection, we present a more difficult dataset with over 1800 annotated human photos.

4. Robert Girshick Donahue, Jeff Toby Darrell Mr. Jitendra Malik. Convolutional networks grounded on regions enabling precise object discovery and segmentation. The competition's last times saw a table in object discovery capability as determined by tests on the sanctioned PASCAL VOC Challenge datasets. Complex ensemble systems with colourful low- position visual attributes and high- position environment were the most effective ways. In this study, we offer a straightforward and scalable discovery algorithm that achieves a mean average perfection (Chart) of 62.4 percent, an increase of further than 50 percent in comparison to the former stylish result on VOC 2012. This. system combines two generalities (1) when labelled training data are scarce, supervised-training for an supplementary task, followed by sphere-specific fine- tuning, significantly improves performance; and (2) when labelled training data are scarce, one can apply high- capacity convolutional networks(CNNs) to bottom- up region proffers in order to localize and member objects. the final model R- CNN or Region- grounded Convolutional Network is related to CNN because the combine region proffers with CNNs.

## III. EXISTING SYSTEM:

In recent times numerous algorithms are developed by numerous experimenters. Both machine literacy and deep literacy approaches work in this operation of computer vision. This section outlines the trip of the different ways used by the experimenters in their study since 2012. SVM algorithm for detecting objects in real time is used. A point sensor which is used to prize meaningful information from the image ignoring the background image. This algorithm works effectively in detecting mortal and textual data. To ameliorate the performance in further general situations. Interesting deep literacy approaches were also used by numerous experimenters in their work. Currently Convolutional Neural Network (CNN) grounded styles were demonstrated to achieve real time object discovery. For e.g. Region of proffers network (RCNN)[5]. RCNN use full image only looks at the portion where the probability of having object is high. It excerpts 2000 regions of every image and ignores rest of the part and takes 45 seconds to reuse every new image. Due to this picky hunt property RCNN works sluggishly and occasionally ignore the important part of the image. After this comes the YOLO family, this another best methods for object detection. RCNN are generally more accurate but YOLO algorithms are important faster and further effective to work in real time discovery. You Only Look Once (YOLO) formerly works on full image by dividing the input image into SXS grid and prognosticating bounding boxes and confidence scores for every grid. Second Version of YOLO algorithm i.e. YOLO_V2 comes with some advancements in terms of perfecting delicacy and making it briskly than YOLO algorithm. YOLO_V2 uses a batch normalization conception which improves the perfection by 2 than original YOLO

algorithm [6]; it also uses a conception of anchor boxes as used in region of proffers system which make YOLO free from all suppositions on bounding boxes. also came the rearmost and the third interpretation of YOLO algorithm (YOLO_V3) which shows slightly better performance and more accurate than YOLO replaces the collective exclusive conception to multi marker bracket i.e. it makes 3 prognostications at each situations. It shows excellent performance in detecting small objects. The main end of this study is object discovery with the ultimate interpretation YOLO_V3 algorithm with audio feedback that can help eyeless peoples to fete all kind of objects near them. As humans can see outside world by using their smarts and eyes and can fluently fete every objects but this capability is lost for visually bloodied peoples[7].

## IV. PROPOSED SYSTEM:

A system that will identify all conceivable daily numerous things and then urge a voice to warn a person about the closest and farthest objects nearby show in in figure 1a[8]

In order to obtain audio at the output of any system for object detection the web speech API is used to create speech at the end which is show as model architecture illustrated in figure 1b

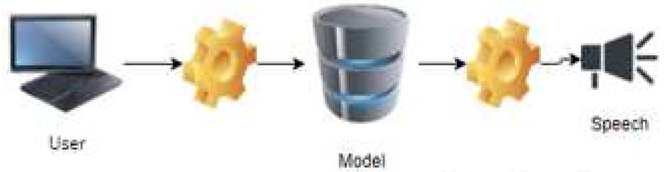


**Fig1a. Model Architecture**

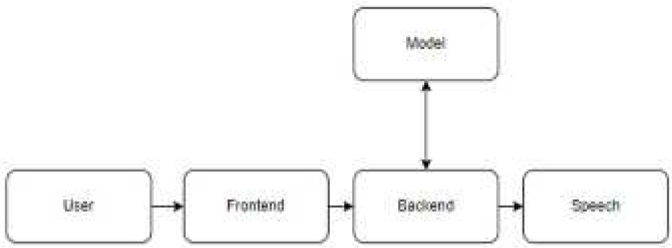


**Fig1b. Model Architecture**

## V. METHODOLOGY:

YOLO-V3 is a part of object detection, Object detection is a computer technology related tocomputer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well- researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance[9-11]. Every object class has its own special features that help in classifying the class. Object class detection uses these special features. For example, when looking for circles, objects that are at a particular distance from a point (i.e., the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed. A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin colour and distance between eyes can be found is shown in figure 2.

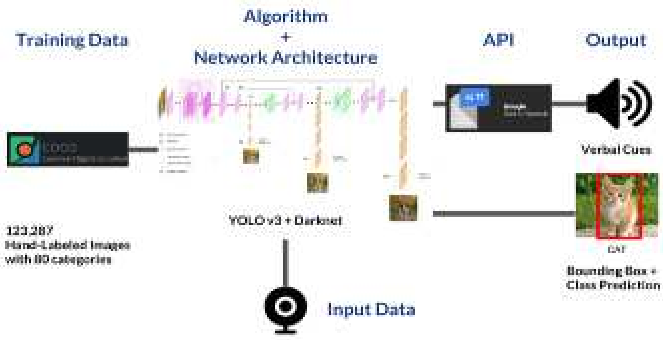


**Fig2: Work flow of YoloV3**

## VI. Results and discussion

From the table 1. it is inferred that the text input is being converted as a speech signal, and image which is being inputted through a camera of fig 2.also available as a speech to recognise that it is a particular object by the visually impaired people.

The object detection of text message and an object is detected using YOLO and YOLO-V3which is shown in figure 3.

Apart from object detection of an image object the performance metrics also measured and comparative analysis had made as a table shown in table 2.

It is obtained from the Table 2 that the precision of YOLO-V3 is superior than YOLO by 10.10%, the Recall also increased by a factor10.08%, the inference time reduced by 1.455sec.Hence, it is claiming that YOLO-V3 is faster than YOLO in detecting the object under test. Fig 2. Is the object detection using YOLO and YOLO-V3 Difference between the two are clearly showing the accuracy and precision.

**Table 1: Results of text and image**

| Input | Output | Result |
|---|---|---|
| Input features | Tested for different features given by user on the model. | Success |
| Images | We were able to detect objectsusing yolov3 and use web speech API to generate speech. | Success |


**Fig3: Difference between YOLO and YOLO-V3**

**Table 2: Performance Metrics**

| Metric in % | YOLO | YOLO-V3 |
|---|---|---|
| precision | 86.44 | 96.50 |
| Recall | 84.90 | 94.98 |
| Inference time | 1.8sec | 0.345sec |

## VII. CONCLUSION:

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well- researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. From the results obtained using YOLO and YOLO-V3 it is concluded that YOLO Family is better than RCNN family.

## REFERENCES:

[1] S. Cherian, & C. Singh, "Real Time Implementation of Object Tracking Through webcam," International Journal of Research in Engineering and Technology, 128-132, (2014)

[2] Z. Zhao, Q. Zheng, P. Xu, S. T, & X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, 30(11), 212-3232, (2019).

[3] N. Dalal, & B. Triggs, "Histograms of oriented gradients for human detection," In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE, (2005, June).

[4] R. Girshick., J. Donahue, T. Darrell, & J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE transactions on pattern analysis and machine intelligence, 38(1), 142-158, (2015).

[5] X. Wang, A. Shrivastava, & A. Gupta, "A-fast-r-cnn: Hard positive generation via adversary for object detection," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2606- 2615), (2017).

[6] S. Ren, K. H, R. Girshick, & J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," In Advances in neural information processing systems (pp. 91-99),(2015).

[7] J. Redmon, S. Divvala, R. Girshick, & A. Farhadi, "You only look once: Unified, real-time object detection," In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788), (2016).

[8] J. Redmon, & A. Farhadi, "YOLO9000: better, faster, stronger," In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271) (2017).

[9] J. Redmon & A. Farhadi, "Yolov3: An incremental improvement," ArXiv preprint arXiv: 1804.02767, (2018).

[10] R. Bharti, K. Bhadane, P. Bhadane, & A. Gadhe, "Object Detection and Recognition for Blind Assistance," International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06, (2019).

[11] T. Lin, Y. Maire, M. Belongie, S. Hays, J. Perona, P. Ramanan, D., & C.L. Zitnick, "Microsoft coco: Common objects in context," In European conference on computer vision (pp. 740-755). Springer, Cham, (2014, September).